# Query2doc: Query Expansion with Large Language Models

## Abstract

This paper introduces a simple yet effective query expansion approach, denoted as query2doc, to improve both sparse and dense retrieval systems. The proposed method first generates pseudo-documents by few-shot prompting large language models (LLMs), and then expands the query with generated pseudo-documents. LLMs are trained on web-scale text corpora and are adept at knowledge memorization. The pseudo-documents from LLMs often contain highly relevant information that can aid in query disambiguation and guide the retrievers. Experimental results demonstrate that query2doc boosts the performance of BM25 by 3% to 15% on ad-hoc IR datasets, such as MS-MARCO and TREC DL, without any model fine-tuning. Furthermore, our method also benefits state-of-the-art dense retrievers in terms of both in-domain and out-of-domain results.